



Otvoreni pristup istraživačkim podacima

Alen Vodopijevec [alen@irb.hr] / Institut Ruđer Bošković, Zagreb

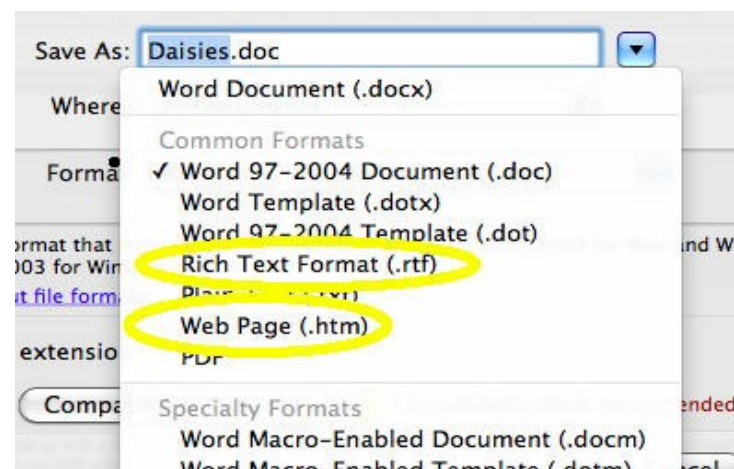


Teme

- **Zašto** istraživački podaci (IP) trebaju biti javno dostupni?
 - općenito o podacima
- **Kako** istraživačke podatke učiniti javno dostupnima?
 - **Planiranje** upravljanja istraživačkim podacima
- **Obveze** u okviru H2020

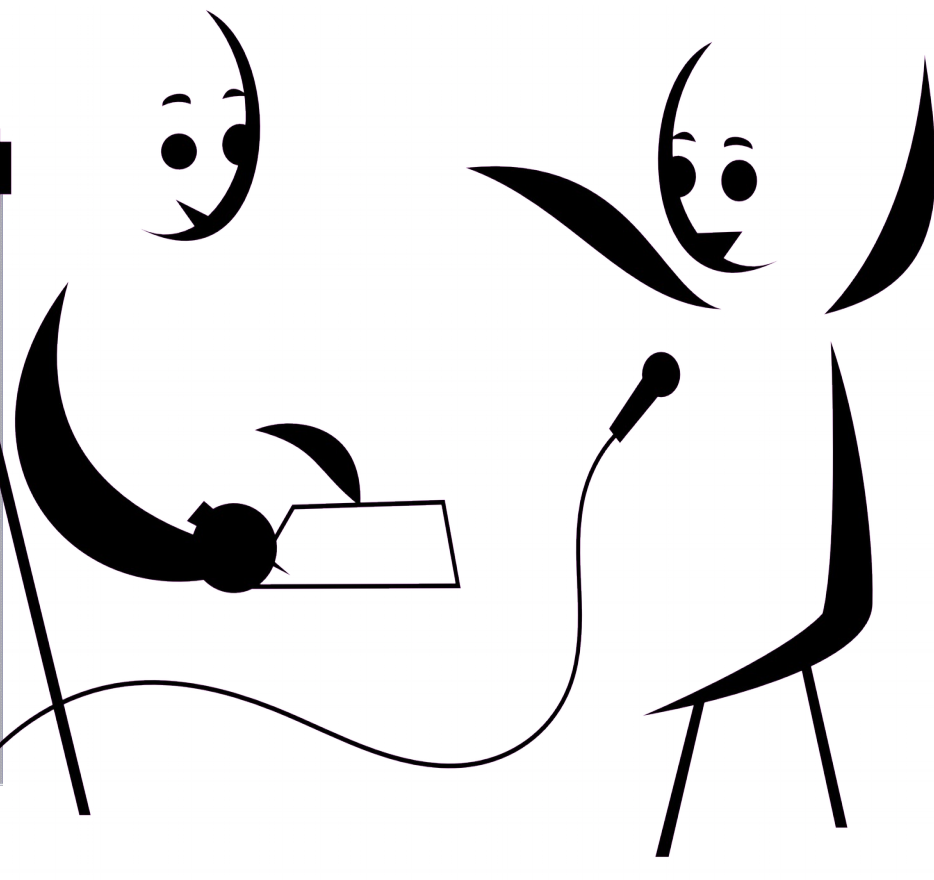
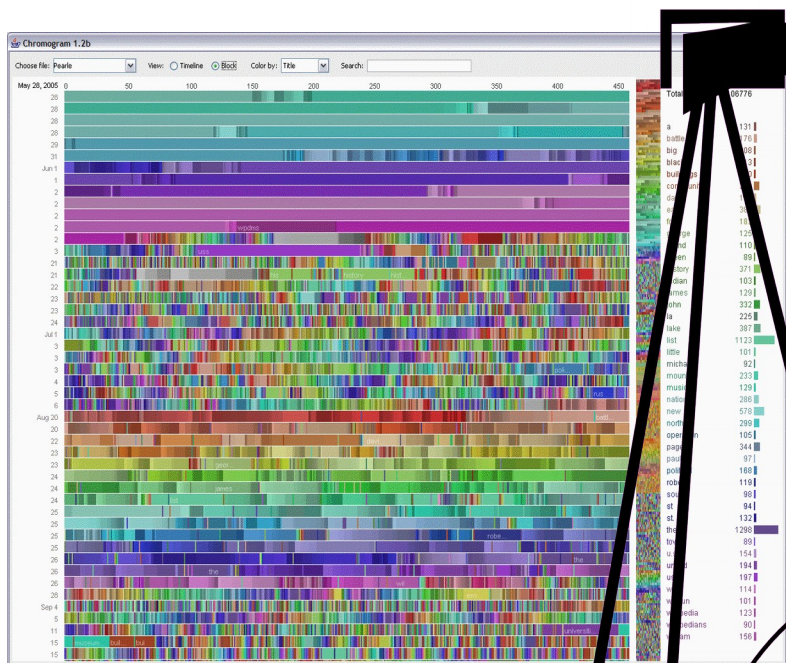
Istraživački podaci (IP)

- IP su podaci prikupljeni, zabilježeni ili generirani u svrhu analize i produciranja originalnih znanstvenih rezultata
- podaci → informacije → znanje
- izvorno digitalno ili digitalizirano



Osnovna podjela IP

- Kvalitativni podaci
- Kvantitativni podaci



Dihotomija nije dovoljna

Podaci mogu biti:

- tekstualni / numerički / multimedijalni
- generirani - modeli / softver
- disciplinarno specifični
- karakteristični za određeni instrument

Imperativ dostupnosti podataka

- Znanost počiva na akumulaciji znanja (s povremenim promjenama paradigmi)
 - Napredak temeljen na postojećim spoznajama
 - Dijeljenje i ponovno korištenje
 - Suradnja
- Da bi se osiguralo **neometano funkcioniranje i “brži” napredak znanosti izvorni istraživački podaci moraju biti slobodno dostupni!***

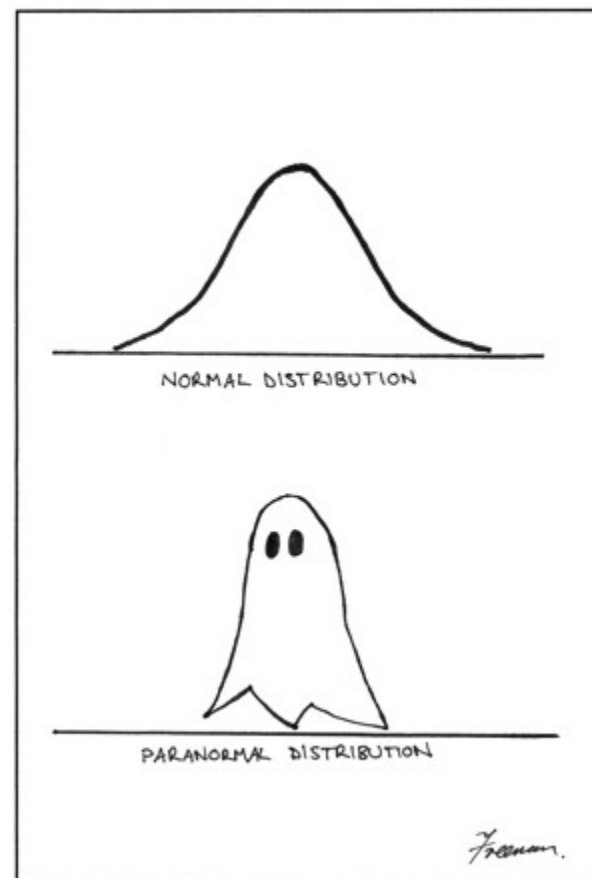
Isti problem jučer i danas

- Prije interneta dostupnost podataka bila je ograničena i vremenskom i prostornom dimenzijom
- Danas postoje svi preduvjeti no bezuvjetno dijeljenje istraživačkih podataka još uvijek nije uobičajena praksa u znanstveno-istraživačkom radu

Pozitivne strane dijeljenja IP

- Validacija rezultata
- Sekundarne analize
- Suradnja
- Obrazovanje

A visual comparison of normal and paranormal distributions, Matthew Freeman, J Epidemiol Community Health 2006;60:6. Lower caption says 'Paranormal Distribution' - no idea why the graphical artifact is occurring



VALIDACIJA REZULTATA

```
int getRandomNumber()  
{  
    return 4; // chosen by fair dice roll.  
              // guaranteed to be random.  
}
```

<http://xkcd.com/221/>

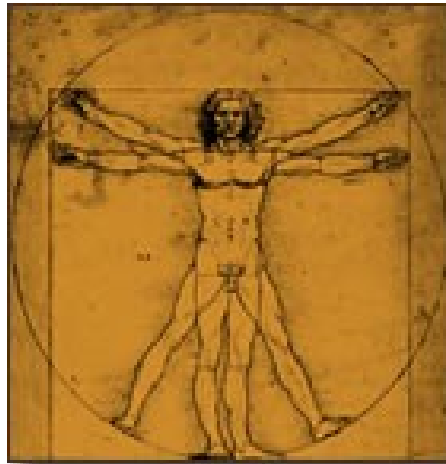
- http://en.wikipedia.org/wiki/Scientific_misconduct
- Razne vrste 'znanstvenih nepodopština'
- slučajevi nizozemskog psihologa Diederika Stapela

SEKUNDARNE ANALIZE

- Ponovno korišćenje postojećih podataka
 - Na drugačiji način
 - Testiranje drugačijih hipoteza
 - Ponavljanje istraživanja te usporedba s novoprikupljenih podataka
 - Meta analize
 - Usporedba s drugim izvorima podataka (?)

SURADNJA

- GENOME projekt



<http://www.genome.gov/10001772>

<http://unlockinglifescode.org/timeline?tid=4>

ftp://ftp.ensembl.org/pub/release-77/fasta/homo_sapiens/dna/

OBRAZOVANJE



http://commons.wikimedia.org/wiki/File:Argonne_lab_education.jpg

OBVEZE SUDIONIKA H2020 PROJEKATA

H2020 Open Data Pilot

- Obveza pohrane i objave istraživačkih podataka nastalih u okviru projekata financiranih sredstvima EK (javna sredstva)

Smjernice za otvoreni pristup publikacijama i podacima u okviru H2020

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

Inicijalno obuhvaćena područja (2014-15)

- Future and Emerging Technologies
- Research infrastructures – part e-Infrastructures
- Leadership in enabling and industrial technologies – Information and Communication Technologies
- Societal Challenge: 'Secure, Clean and Efficient Energy' – part Smart cities and communities
- Societal Challenge: 'Climate Action, Environment, Resource Efficiency and Raw materials' – except raw materials
- Societal Challenge: 'Europe in a changing world – inclusive, innovative and reflective Societies'
- Science with and for Society

H2020 Open Data Pilot

- Izrada Plana za upravljanje istraživačkim podacima tijekom prvih 6 mjeseci trajanja projekta
 - *DMPOnline – Alat za kreiranje plana*
- Mogućnost izuzeća od te obveze (opt out)
- Objava izvornih podataka potrebnih za validaciju usporedo s objavom rada
- Omogućavanje besplatnog pristupa podacima korištenjem CC-BY ili CC0 licenci

KAKO i GDJE OBJAVITI PODATKE?

- Kada ne postoji specijalizirani arhiv?
- Razmisliti o:
 - Vidljivosti i dostupnosti pohranjenih podataka
 - Reguliranju prava pristupa
 - Razumljivosti i pristupačnosti
 - Prezervaciji

VIDLJIVOST

- Kako će potencijalni korisnici pronaći konkretne podatke?
 - Institucijski repozitorij
 - [ZENODO](#) (OpenAIRE projekt)
 - Katalozi tematskih repozitorija
 - [Databib](#)
 - [Re3data.org](#)
 - Repozitorij časopisa
- Trajni/postojani identifikatori

PRAVA PRISTUPA i PRAVILA KORIŠTENJA

- **Tko, kada i pod kojim uvjetima može pristupiti podacima te na koji način ih smije koristiti?**
- **Panton Principles**
 - Korištenje što otvorenijih licenci
 - <http://pantonprinciples.org/>
- **Definicija otvorenog znanja i usporedba licenci, [infografika]**
- **Pravna i etička pitanja!**

RAZUMLJIVOST i PRISTUPAČNOST

- Kako i na koji način će podaci biti opisani
 - Dokumentacija
 - Kontekst
- Metapodaci
 - Standardi metapodataka često ovise o disciplini ili o specifičnom znanstvenom području interesa

PREZERVACIJA

- Sigurnost pohranjenih podataka
- Osiguravanje dugoročne pohrane
- Planiranje mogućnosti pristupa i korištenja u budućnosti
 - Formati podataka, migracija
- Potencijalni troškovi prezervacije
 - Problem licenciranja CC NC ND

STANJE INFRASTRUKTURE U RH

Arhiviranje i životni ciklus podataka

- **Generiranje** podataka - osmišljavanje istraživanja, plana upravljanja istraživačkim podacima, obrasci pristanka na dijeljenje, identifikacija postojećih dostupnih podataka, prikupljanje (eksperiment, opažanje, mjerenje, simulacija)
- **Procesiranje** podataka – priprema za analizu
- **Analiza** podataka
- **Prezervacija** podataka - pohrana i sigurnose kopije, arhiviranje, opisivanje metapodacima, migracija u formate prikladne za dugotrajno očuvanje
- **Osiguravanje pristupa** podacima - dijeljenje podataka, kontrola pristupa
- **Ponovno korištenje** podataka – re-use

Plan razvoja istraživačke i inovacijske infrastrukture u Republici Hrvatskoj

U sljedećem razdoblju potrebno je osigurati:

- *g) uspostavu i sustavni razvoj podatkovnog sloja e-infrastrukture koji treba omogućiti sustave i alate za razmjenu, zajedničko korištenje i čuvanje istraživačkih podataka ...*
- *i) poticanje otvorenog pristupa istraživačkim podacima, obavezno podacima koji su nastali financiranjem iz javnih izvora;*

<http://public.mzos.hr/Default.aspx?art=13111> MZOS, 1. travanj 2014.

Projektne aktivnosti u području upravljanja istraživačkim podacima i otvorenog pristupa

- JKP, SERSCIDA prototip, OpenAIREPlus
- Sudjelovanja u H2020 projektima
 - OpenAIRE2020 (početak 1.2015.) → cris
 - CESSDA SaW (priprema prijedloga)
- SCOPES
 - SEEDS (evaluacija u tijeku)
- FOSTER
 - evaluacija prijava drugog kruga u tijeku

Zaključak

- Dijeliti se mora
 - Otvoreni pristup IP je ključan za brži razvoj znanosti
 - Obveze pohrane i objave IP
- Razvoj nacionalne infrastrukture
 - Podrška procesima upravljanja istraživačkim podacima
 - *Backoffice / Frontoffice* paradigma

Rasprava